

Patient Specific QA in Adaptive Radiotherapy (ART)

Options, Limitations, & Future Directions

Justus Adamson, PhD

Associate Professor

Duke Radiation Oncology



DukeHealth





Conflict of Interest Statement

- Research funding from Radialogica LLC
- Ownership of Clearsight RT LLC (unrelated to this topic)



LIVE

breakyourownnews.com

BREAKING NEWS

ADAPTIVE RT BREAKS QA PARADIGM

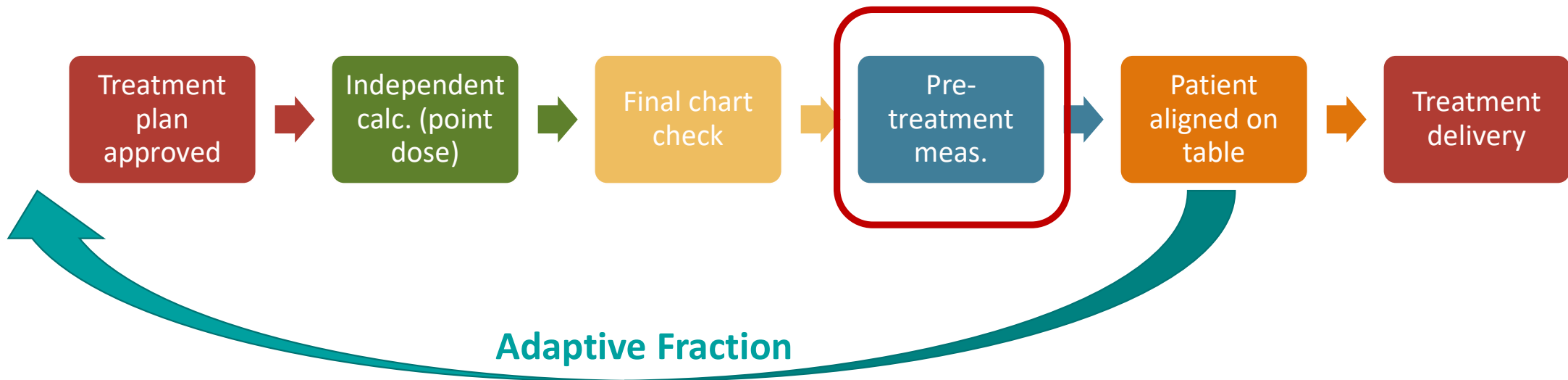
11:04

A NEW RADIOTHERAPY TREATMENT TECHNIQUE KNOWN AS "ADAPTIVE RADIOTHER



Radiotherapy Flowchart:

How to perform a pre-treatment measurement with the patient on the table?





Patient Specific QA in Adaptive Radiotherapy:

Med. Phys. 41 (8), August 2014

Process-based quality management for clinical implementation of adaptive radiotherapy

Camille E. Noel, Lakshmi Santanam, Parag J. Parikh, and Sasa Mutic^{a)}
Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri 63110

(Received 25 September 2013; revised 22 May 2014; accepted for publication 3 July 2014;
published 30 July 2014)

Purpose: Intensity-modulated adaptive radiotherapy (ART) has been the focus of considerable research and developmental work due to its potential therapeutic benefits. However, in light of its unique quality assurance (QA) challenges, no one has described a robust framework for its clinical implementation. In fact, recent position papers by ASTRO and AAPM have firmly endorsed pretreatment patient-specific IMRT QA, which limits the feasibility of online ART. The authors aim to address these obstacles by applying failure mode and effects analysis (FMEA) to identify high-priority errors and appropriate risk-mitigation strategies for clinical implementation of intensity-modulated ART.

Methods: An experienced team of two clinical medical physicists, one clinical engineer, and one radiation oncologist was assembled to perform a standard FMEA for intensity-modulated ART. A set of 216 potential radiotherapy failures composed by the forthcoming AAPM task group 100 (TG-100) was used as the basis. Of the 216 failures, 127 were identified as most relevant to an ART scheme. Using the associated TG-100 FMEA values as a baseline, the team considered how the likeliness of occurrence (O), outcome severity (S), and likeliness of failure being undetected (D) would change for ART. New risk priority numbers (RPN) were calculated. Failures characterized by $RPN \geq 200$ were identified as potentially critical.

Results: FMEA revealed that ART RPN increased for 38% ($n = 48/127$) of potential failures, with 75% ($n = 36/48$) attributed to failures in the segmentation and treatment planning processes. Forty-three of 127 failures were identified as potentially critical. Risk-mitigation strategies include implementing a suite of quality control and decision support software, specialty QA software/hardware tools, and an increase in specially trained personnel.

Conclusions: Results of the FMEA-based risk assessment demonstrate that intensity-modulated ART introduces different (but not necessarily more) risks than standard IMRT and may be safely implemented with the proper mitigations. © 2014 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4890589>]

FMEA of ART assuming no IMRT QA

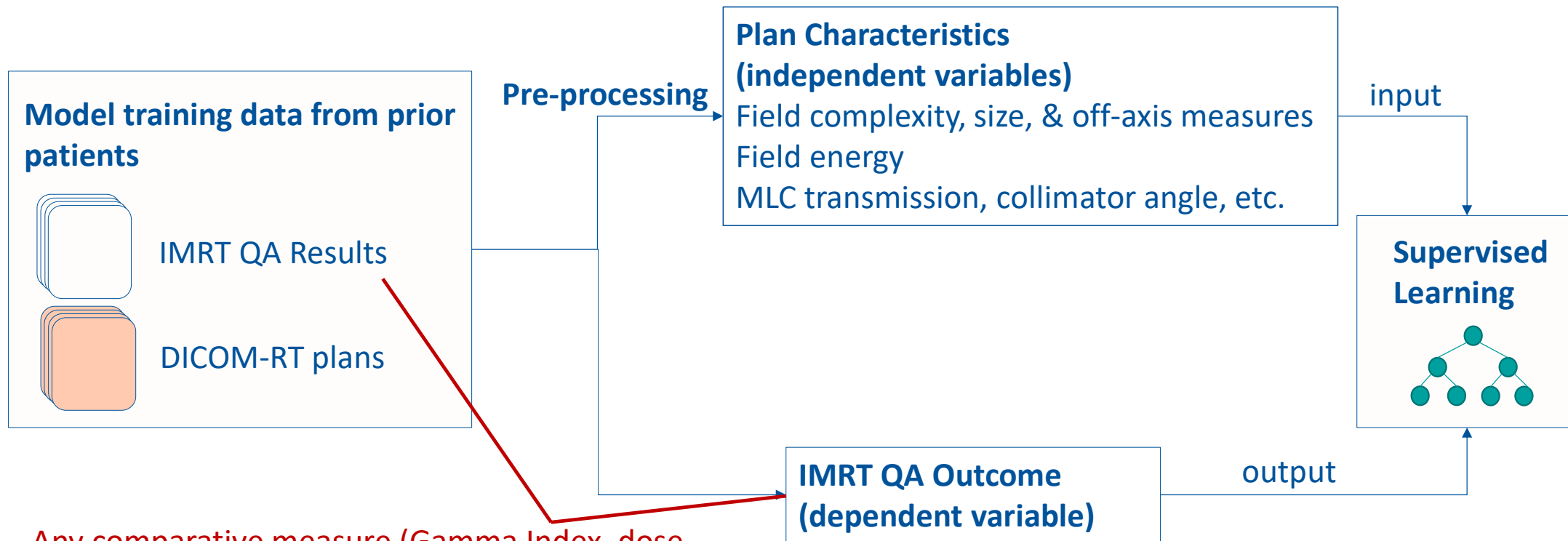
- Concluded that ART introduces different (but not necessarily more) risks than standard IMRT
- Proposed Alternatives to IMRT-QA
 - Quality Assurance with Plan Veto (QAPV) software
 - Simulated delivery (dry-run) with patient on table
 - Post treatment or real-time monitoring of log files
 - EPID based in-vivo dosimetry
 - Transmission detectors



Our Focus:
AI prediction models for ART
patient specific QA



AI Prediction Models for IMRT QA: General Concept



- Any comparative measure (Gamma Index, dose difference, etc.)
- Any QA measurement device (EPID, ion chamber & diode arrays, etc.)



AI Prediction Models for IMRT QA:

Interian *et al.*: Deep learning vs medical physicists

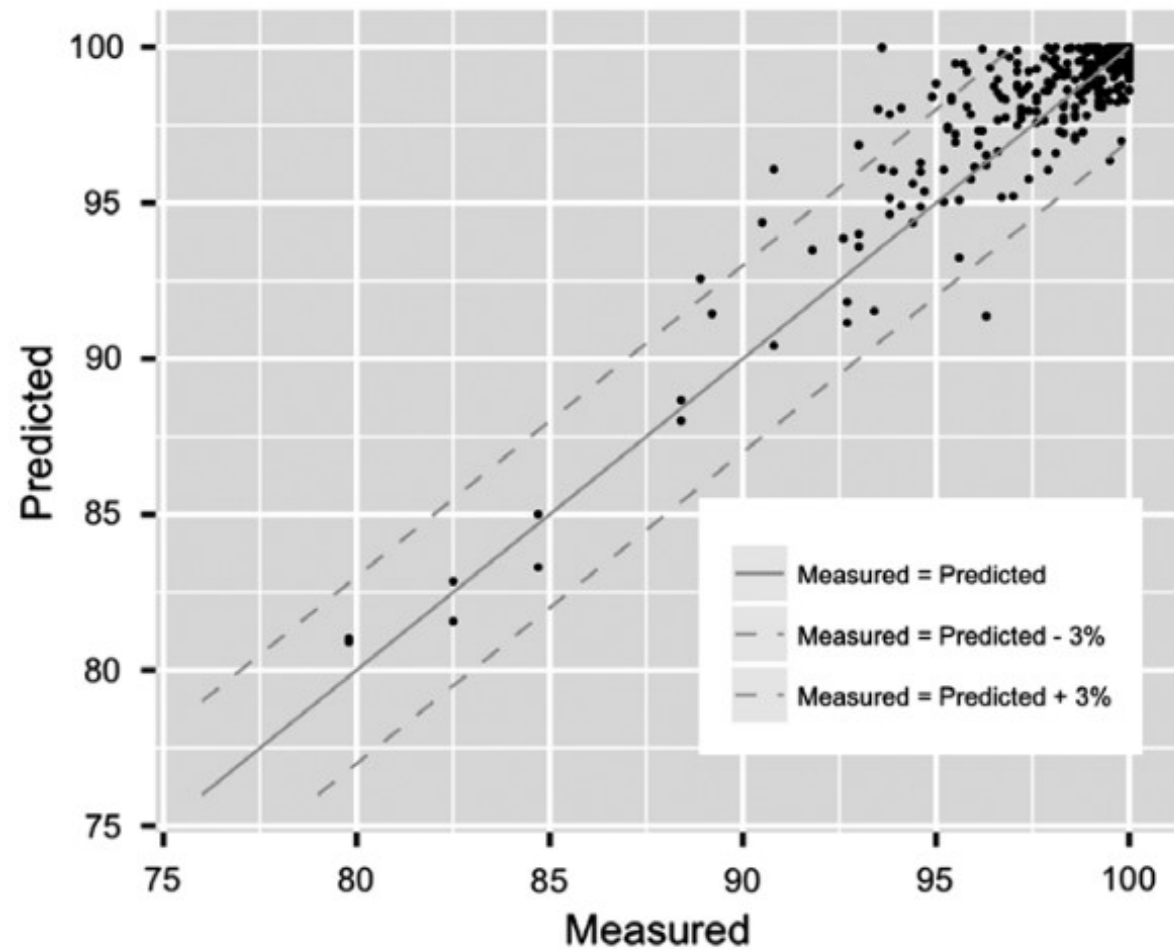
Medical Physics, 45 (6), June 2018

492 Mapcheck2 measurements



(a)

Measured vs Predicted Passing Rate (Neural Network)





Background:

A mathematical framework for virtual IMRT QA using machine learning

G. Valdes,^{a),b)} R. Scheuermann,^{a)} C. Y. Hung, A. Olszanski, M. Bellerive, and T. D. Solberg
Radiation Oncology Department, Perelman Center for Advanced Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19123

(Received 16 March 2016; revised 31 May 2016; accepted for publication 1 June 2016;
published 20 June 2016)

Deep nets vs expert designed features in medical physics: An IMRT QA case study

Yannet Interian* and Vincent Rideout*

MS in Analytics Program, University of San Francisco, San Francisco, CA, USA

Vasant P. Kearney, Efsthios Gennatas, Olivier Morin, Joey Cheung, Timothy Solberg, and Gilmer Valdes^{a)}

Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA

(Received 6 October 2017; revised 9 January 2018; accepted for publication 9 January 2018;
published 18 April 2018)

RADIATION ONCOLOGY PHYSICS

IMRT QA using machine learning: A multi-institutional validation

Gilmer Valdes^{1,3,a} | Maria F. Chan^{2,a} | Seng Boh Lim² | Ryan Scheuermann³ |
O. Deasy² | Timothy D. Solberg^{1,3}

Physics Contribution

Machine Learning for Patient-Specific Quality Assurance of VMAT: Prediction and Classification Accuracy

Jiaqi Li, MS,* Le Wang, PhD,^{†,‡} Xile Zhang, MS,* Lu Liu, MS,*
Jun Li, PhD,* Maria F. Chan, PhD,[§] Jing Sui, PhD,^{†,‡}
and Ruijie Yang, PhD*

Tomohiro Ono^{a)}, Hideaki Hirashima, Hiraku Iramina, Nobutaka Mukumoto, and Yuki Miyabe
Department of Radiation Oncology and Image-applied Therapy, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan

Mitsuhiro Nakamura
Department of Radiation Oncology and Image-applied Therapy, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan
Division of Medical Physics, Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School Medicine, Kyoto University, Kyoto, Japan

Takashi Mizowaki
Department of Radiation Oncology and Image-applied Therapy, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan

Application and comparison of machine learning models for predicting quality assurance outcomes in radiation therapy treatment planning

Phillip D.H. Wall^{a,*}, Jonas D. Fontenot^{a,b}

^a Department of Physics and Astronomy, Louisiana State University and Agricultural and Mechanical College, 202 Tower Drive, Baton Rouge, LA, 70803-4001, USA

^b Department of Physics, Mary Bird Perkins Cancer Center, 4950 Essen Lane, Baton Rouge, LA, 70809, USA

Predicting VMAT patient-specific QA results using a support vector classifier trained on treatment plan characteristics and linac QC metrics

Dal A Granville^{1,4}, Justin G Sutherland^{1,2,3}, Jason G Belec^{1,2} and Daniel J La Russa^{1,2,3,4}

¹ Radiation Medicine Program, The Ottawa Hospital, Ottawa, Canada

² Division of Radiation Oncology, Department of Radiology, The University of Ottawa, Ottawa, Canada

³ The Ottawa Hospital Research Institute, Ottawa, Canada

⁴ Author to whom any correspondence should be addressed.

E-mail: dgranville@toh.ca

RESEARCH

Open Access

Prediction of VMAT delivery accuracy with textural features calculated from fluence maps

Jong Min Park^{1,2,3,4}, Jung-in Kim^{1,2,3} and So-Yeon Park^{2,5*}

Predicting gamma passing rates for portal dosimetry-based IMRT QA using machine learning

Dao Lam

Department of Radiation Oncology, Washington University School of Medicine, 4921 Parkview Place, Campus Box 8224, St. Louis, MO 63110, USA

Xizhe Zhang

School of Computer Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China

Harold Li, Yang Deshan, Brayden Schott, and Tianyu Zhao

Department of Radiation Oncology, Washington University School of Medicine, 4921 Parkview Place, Campus Box 8224, St. Louis, MO 63110, USA

Weixiong Zhang

Department of Computer Science and Engineering, Washington University, One Brookings Drive, Campus Box 1045, St. Louis, MO 63130, USA

Sasa Mutic, and Baozhou Sun^{a)}

Department of Radiation Oncology, Washington University School of Medicine, 4921 Parkview Place, Campus Box 8224, St. Louis, MO 63110, USA

Lot's of examples & counting!



AI prediction of IMRT QA Results: Limitations

- In ideal case, AI model *perfectly* reflects outcome from IMRT QA measurement
- **Inherent limitations** from measurement device **will be reflected** in the AI model
 - **False positives** (examples: poor EPID scatter modeling causes failures off axis)
 - **False negatives** (examples: limited detector sensitivity, resolution, lack of comprehensive measurement etc.)

An effective correction algorithm for off-axis portal dosimetry errors

Daniel W. Bailey

Department of Physics, State University of New York at Buffalo, Buffalo, New York 14260

Lalith Kumaraswamy and Matthew B. Podgorsak

Department of Radiation Medicine, Roswell Park Cancer Institute, Buffalo, New York 14263

(Received 19 November 2008; revised 1 July 2009; accepted for publication 2 July 2009; published 12 August 2009)

Portal dosimetric images acquired for IMRT pretreatment verification show dose errors of up to 15% near the detector edges as compared to dose predictions calculated by a treatment planning system for these off-axis regions. A method is proposed to account for these off-axis effects by precisely correcting the off-axis output factors, which calibrate the imager for absolute dose. Using this method, agreement between the predicted and the measured doses improves by up to 15% for fields near the detector edges, resulting in passing rate improvements of as much as 60% for gamma evaluation of 3 mm, 3% within the collimator jaws. © 2009 American Association of Physicists in Medicine. [DOI: [10.1118/1.3187785](https://doi.org/10.1118/1.3187785)]

Institutional Patient-specific IMRT QA Does Not Predict Unacceptable Plan Delivery

Kry et al. Volume 90 • Number 5 • 2014

International Journal of Radiation Oncology • Biology • Physics

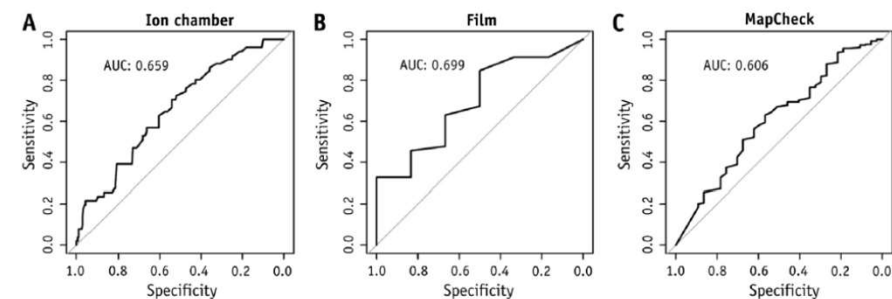


Fig. 5. ROC curves for the ion chamber (a), film (b), and MapCheck (c) devices, indicating the sensitivity and specificity of the devices across thresholds (dose difference for the ion chamber and percent of pixels passing 3%/3-mm gamma for film and MapCheck). The AUC is shown for each device. AUC = area under the curve; ROC = receiver operator characteristic.



AI prediction of IMRT QA Results: Limitations

Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors^{a)}

Med. Phys. 38 (2), February 2011

Conclusions: There is a lack of correlation between conventional IMRT QA performance metrics (Gamma passing rates) and dose errors in anatomic regions-of-interest. The most common acceptance criteria and published actions levels therefore have insufficient, or at least unproven, predictive power for per-patient IMRT QA. © 2011 American Association of Physicists in Medicine

Pretreatment patient-specific IMRT quality assurance: A correlation study between gamma index and patient clinical dose volume histogram

Γ -index pass rates are of limited value, even if predicted perfectly.

only weak correlations (Pearson's r -values) were observed, including some false negatives (cases where high IMRT QA performance metrics and the reverse, mainly for the

Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA

Medical Physics, Vol. 38, No. 10, October 2011

Conclusions: Gamma passing rate, even if calculated based on patient dose grids, has generally weak correlation to critical patient DVH errors. However, the PDP algorithm was shown to accurately predict the DVH impact using conventional planar QA results. Using patient-DVH-based

IMRT QA performance metrics gamma passing rates and dose errors in DVHs values and the low sensitivity of 3%/3 mm global gamma method show that the most common published acceptance criteria have disputable predictive power for per-patient IMRT QA. © 2012 American Association of Physicists in Medicine.

metrics IMRT QA allows per-patient dose specific. Further studies are now required with DVH-based metrics to ensure effective

AI model does not enable translation of Γ -index pass rates to clinically meaningful measures (patient DVH)



Alternative: Predict patient dose effect *directly*

Proposal

- Predict actual machine parameters at treatment delivery
 - **Input data:** DICOM-RT plan
 - **Model training data:** DICOM-RT plans & log files from prior patients
- Use this prediction for pre-treatment dose calculation on patient anatomy

Potential Advantages

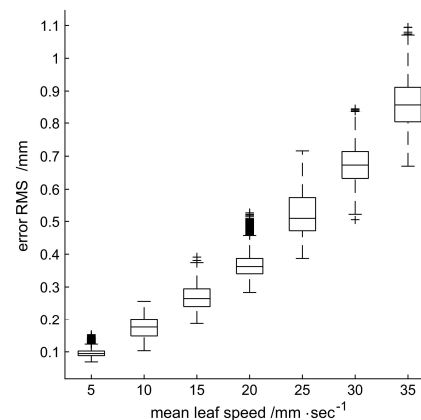
- Patient specific prediction
- Predict a meaningful result
 - Dose distribution / DVH
- Low maintenance
 - Automatically acquired training / validation data
- Enable a virtual process
 - Applicable to adaptive techniques!



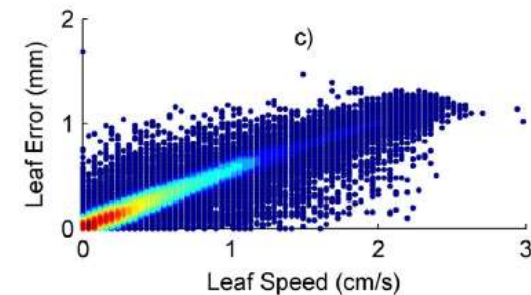
Feasibility of predicting log file(s)

- Difference between planned MLC positions & actual at treatment delivery are very predictable
- High correlation with MLC velocity

Kabat *et al. Medical Physics*,
46 (3), March 2019

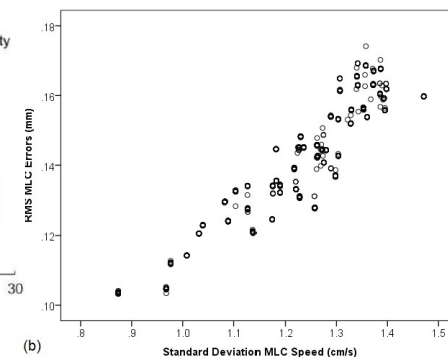
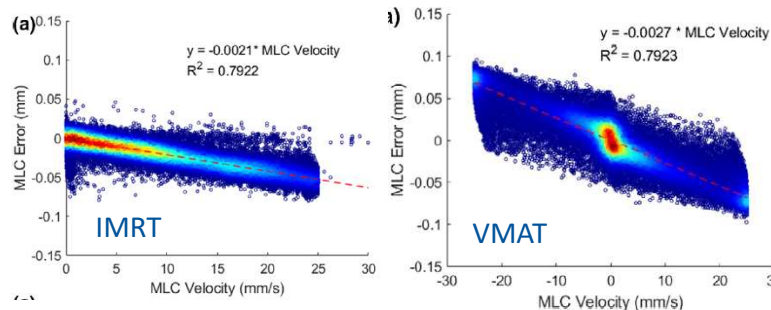


J. Olasolo-Alonso *et al. Physica Medica* 33 (2017) 87–94

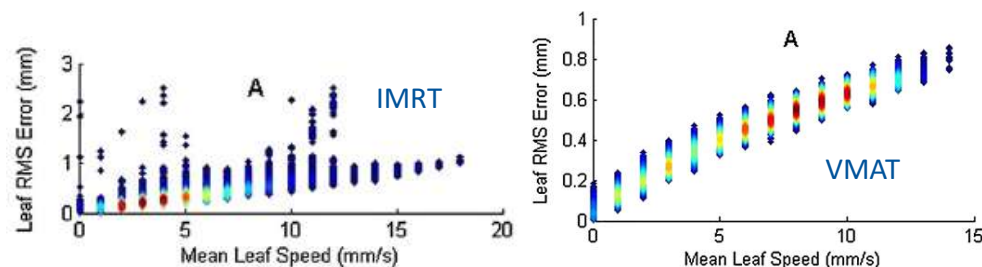


Agnew *et al. JACMP* Vol.
15, No. 6, 2014

Chuang *et al., Medical Physics*, 48 (3), March 2021



Kerns *et al. Radiation Oncology* 2014, 9:176



AI Prediction of Log Files: Relevant Prior Work

Carlson et al. Phys. Med. Biol. 61 25, 2016

- 74 VMAT plans from 3 institutions
- Training (data from 1 plan)
- Models: Linear regression, Random forest, Cubist

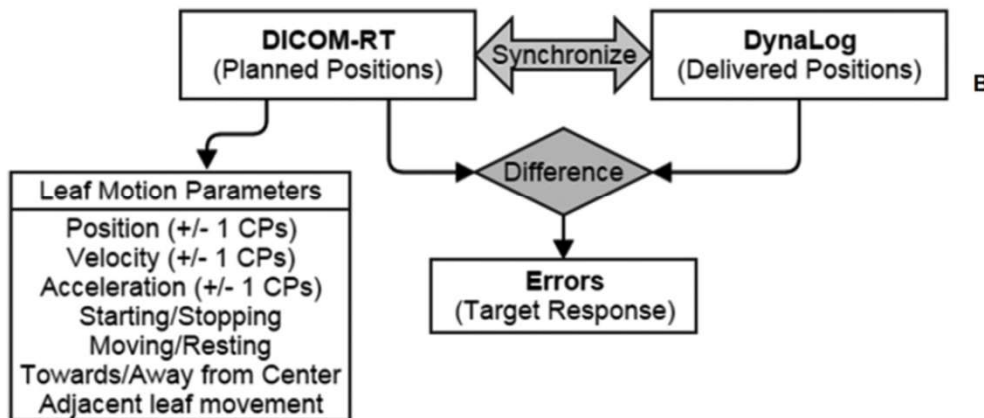


Figure 1. Workflow of the extraction of errors between DICOM-RT and DynaLog files, and the extraction of leaf motion parameters from planned positions.

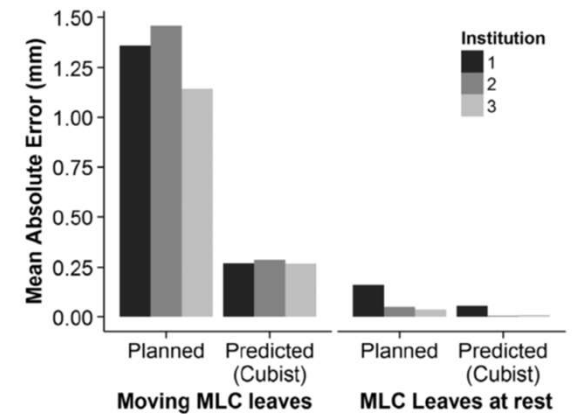
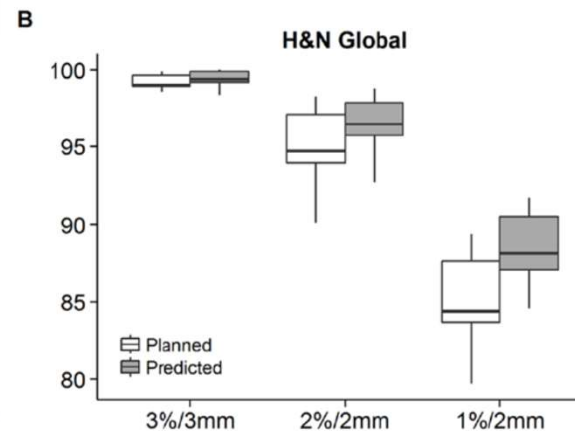
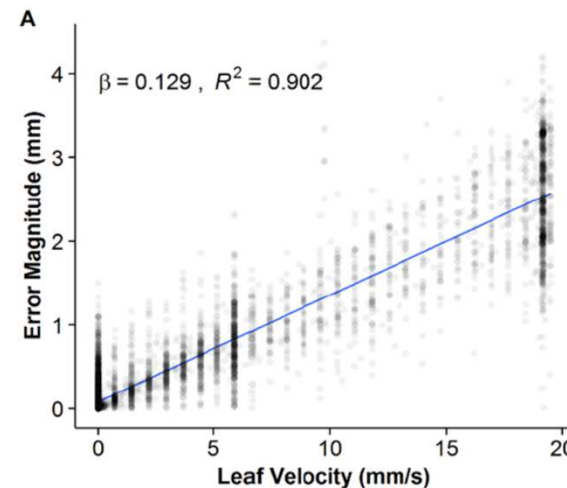
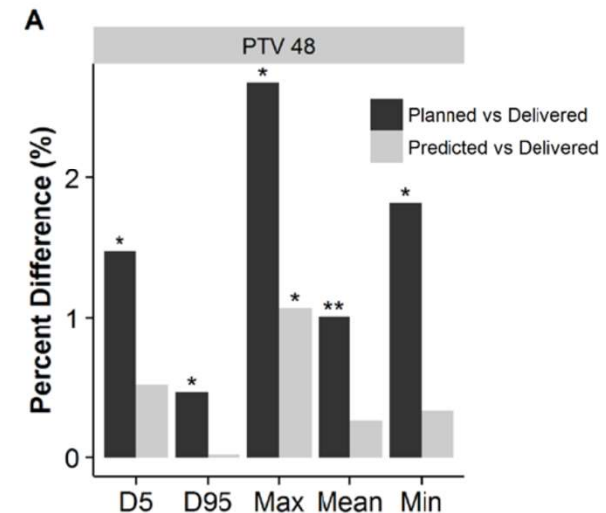
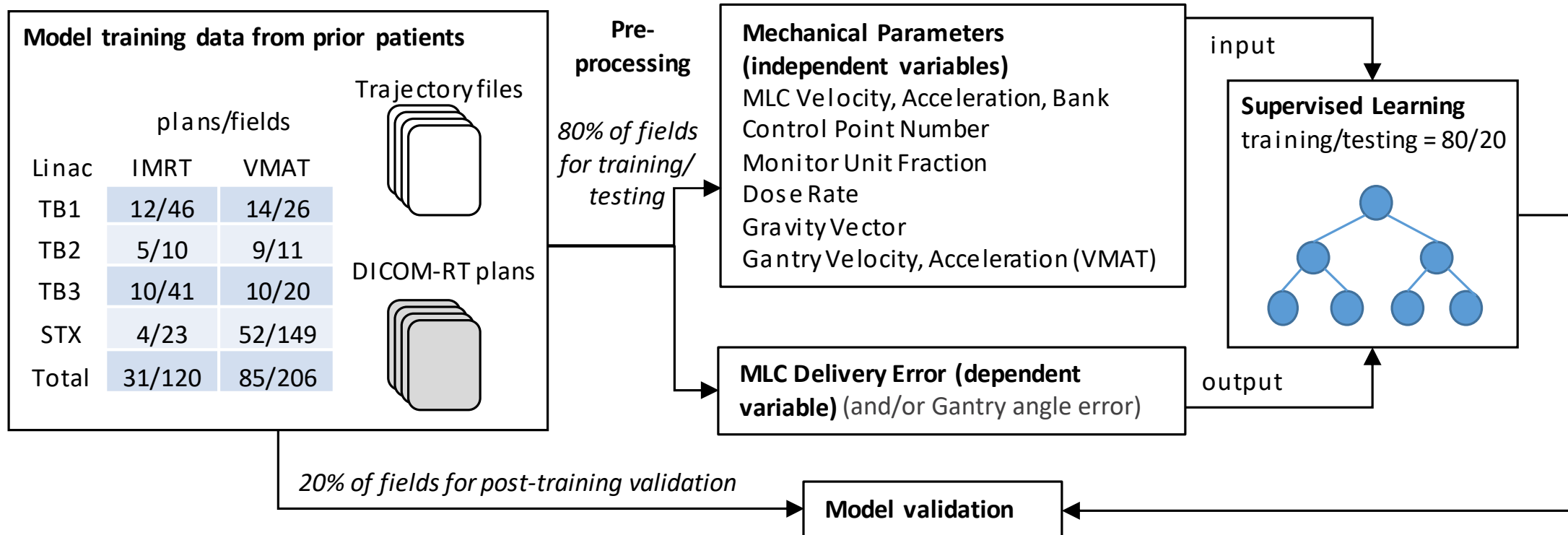


Figure 5. MAE between planned and delivered positions, and between predicted and delivered positions for both resting and moving MLC leaves. In all cases the predicted positions are much closer to the delivered positions than are the planned positions.



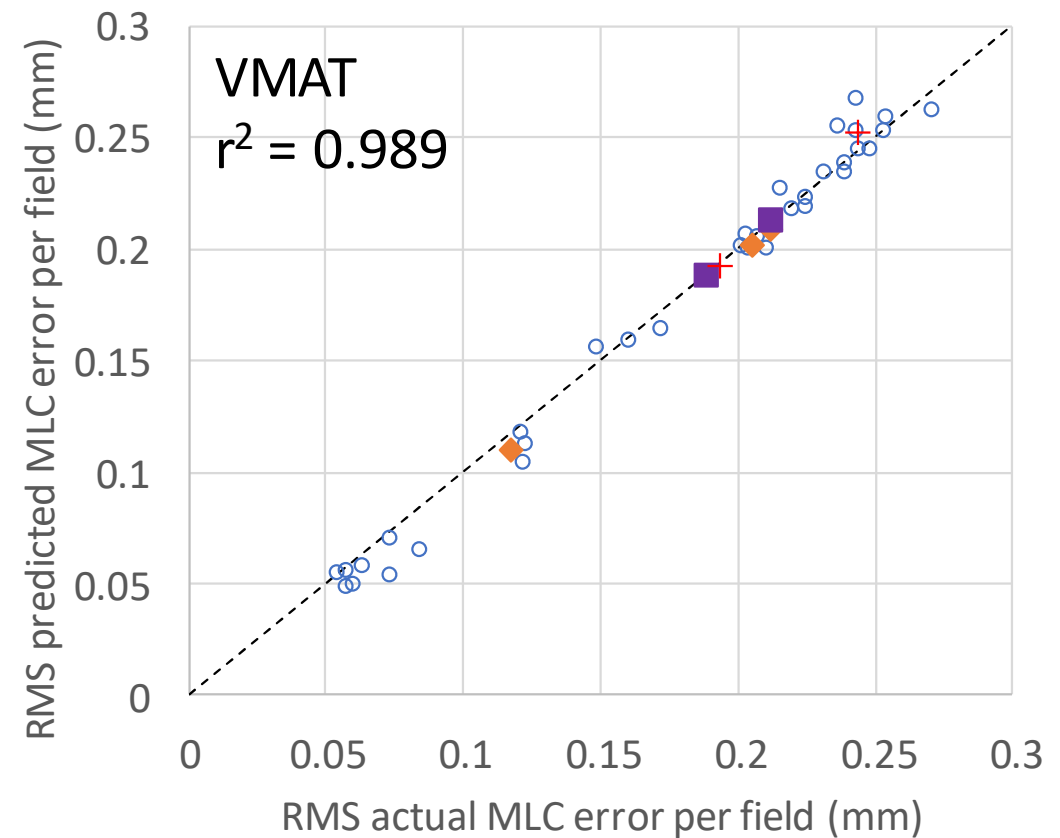
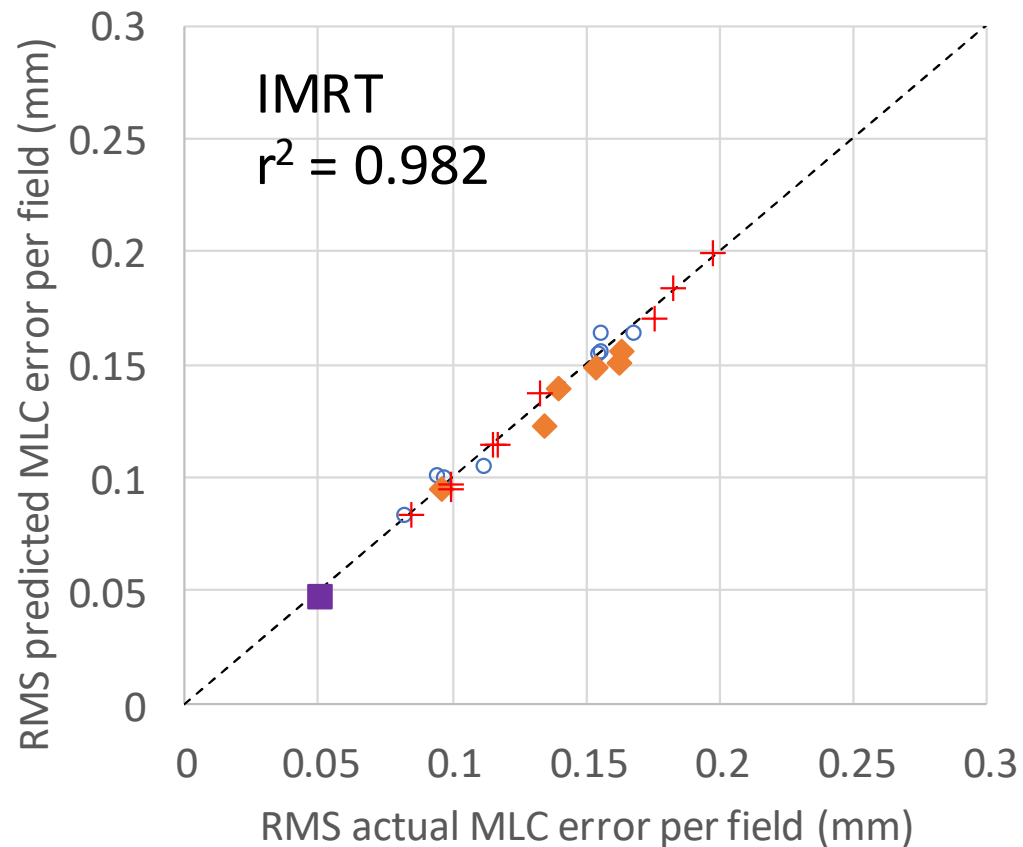


Current Work: Model Training, Testing, and Validation





Model Performance: MLC discrepancy



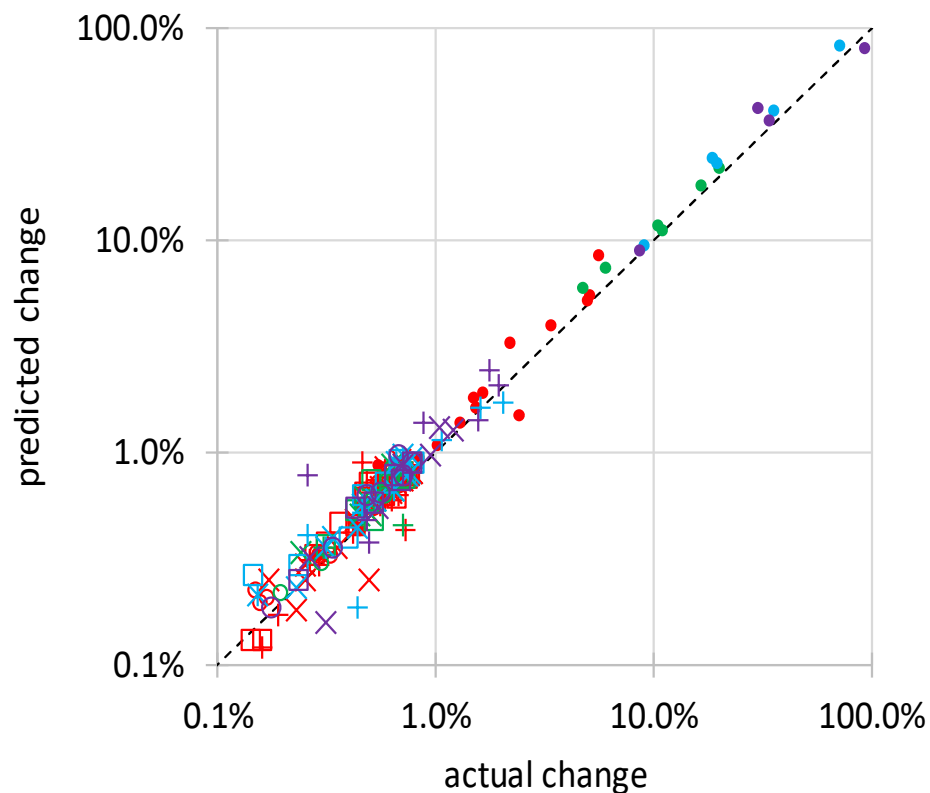
○ STX ◆ TB1 ■ TB2 + TB3 - - - - y=x

Lay, et al. "Virtual patient-specific QA with DVH-based metrics." *JACMP* (2022).

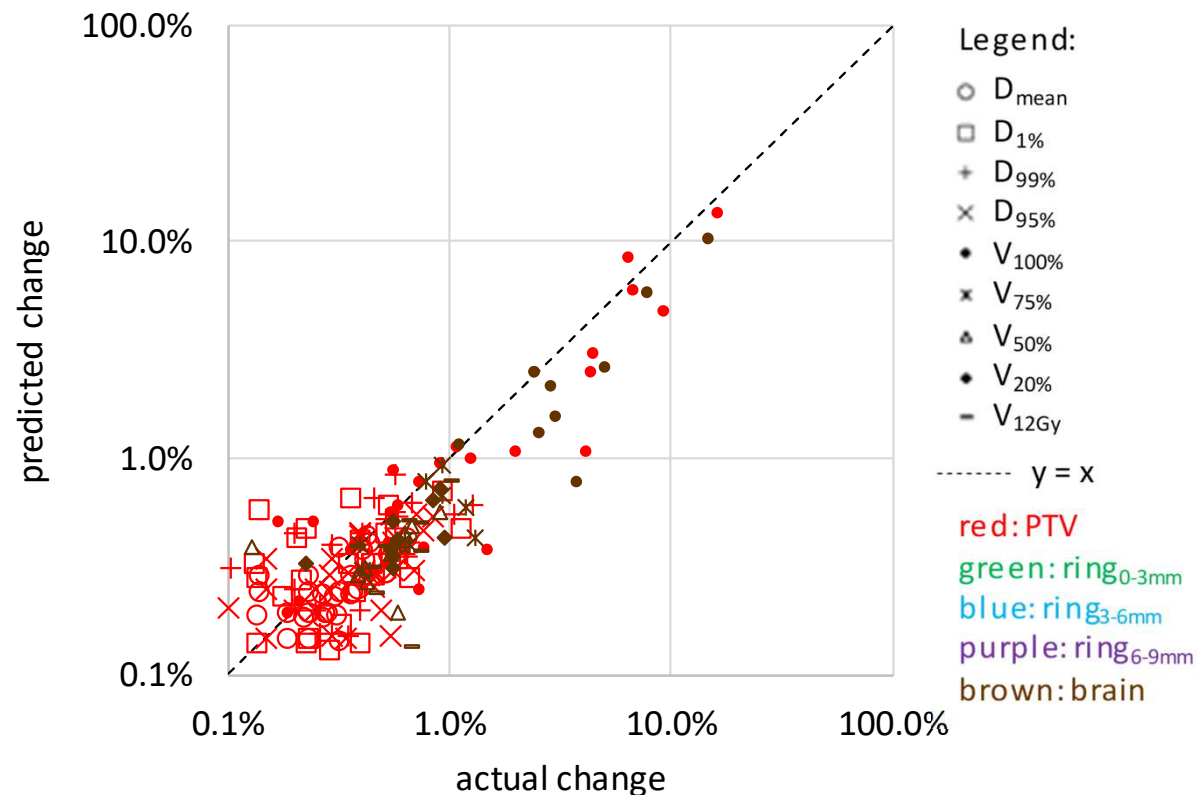


Model Performance: (Dosimetric)

IMRT



VMAT

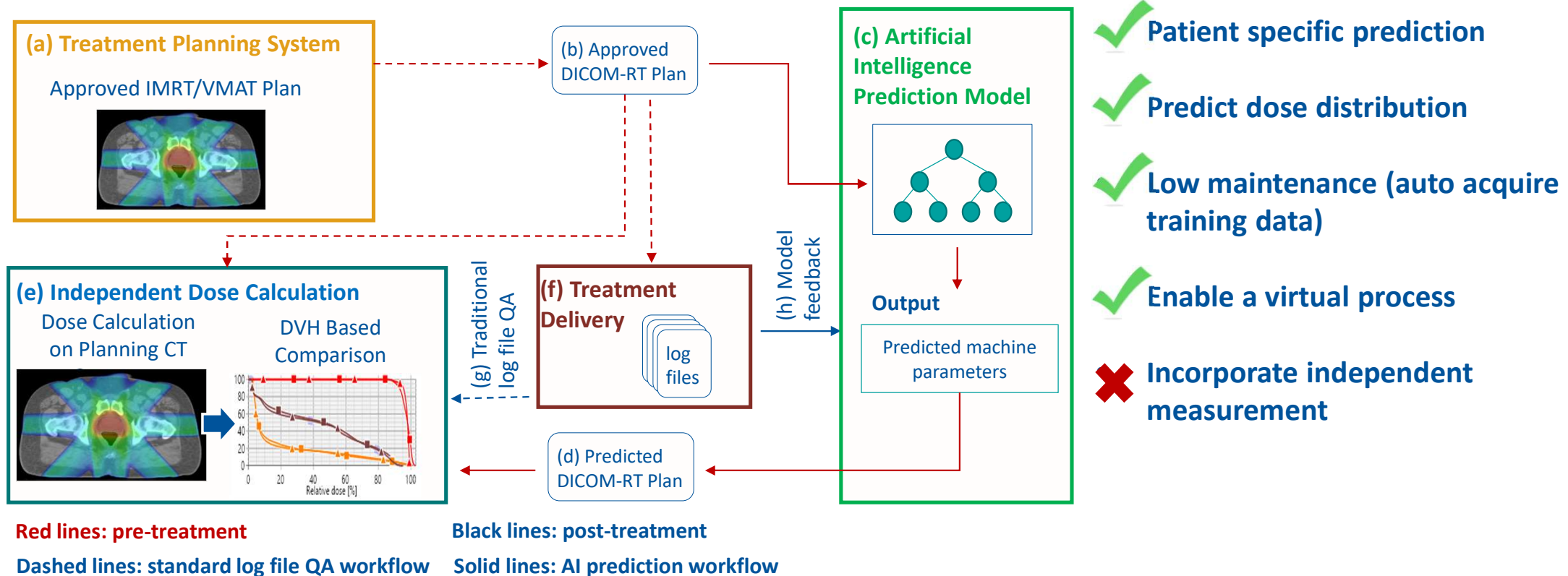


Lay, et al. "Virtual patient-specific QA with DVH-based metrics." *JACMP* (2022).



How would this work?

Step 1: Log file based virtual QA:



Lay, et al. "Virtual patient-specific QA with DVH-based metrics." *JACMP* (2022).

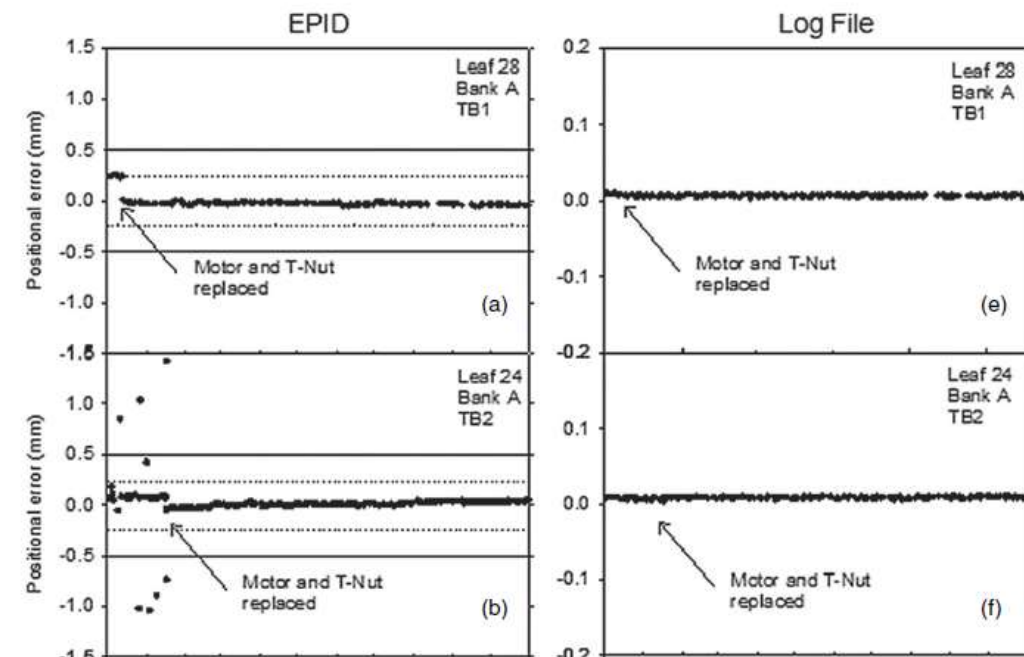
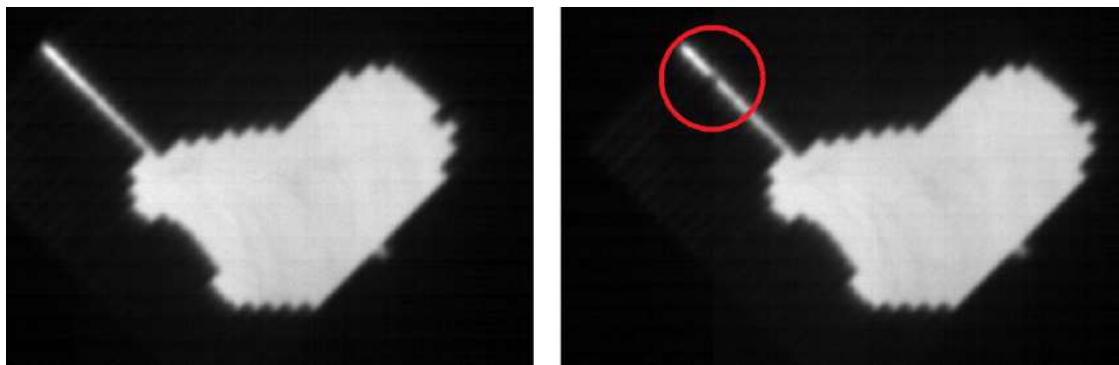


Log file prediction model: Limitations

Limitation 1: Log files do not see all MLC errors

Neal *et al*: *Med. Phys.*, Vol. 43, No. 6

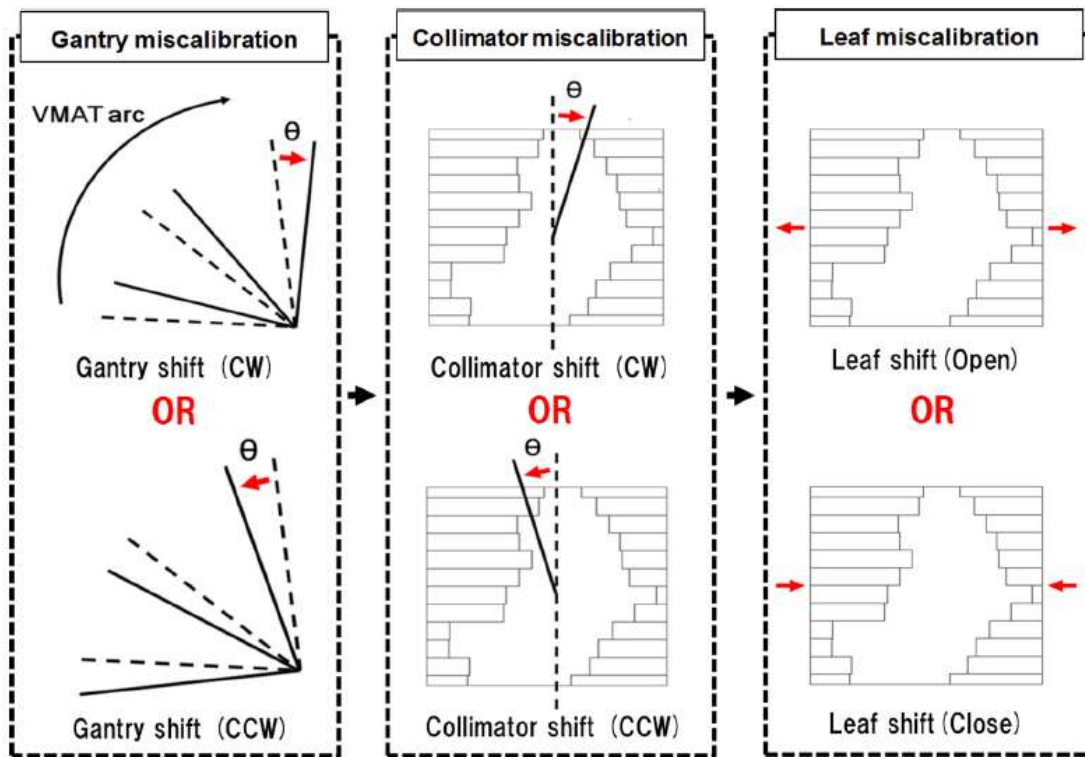
A Agnew *et al*: *PMB* 59 (2014) N49





Log file prediction model: Limitations

Limitation 2: Log file assumes machine is calibrated perfectly
(does not account for residual miscalibration errors)



Dosimetric effect of mis-calibration errors:

Treatment site	Structure	General tolerance level		
		D_{mean} (Gy)	$D_{2\%}$ (Gy)	$D_{98\%}$ (Gy)
Head-and-neck	PTV	0.81 ± 0.15	0.77 ± 0.28	0.86 ± 0.32
	Parotid	0.81 ± 0.61	0.89 ± 0.53	0.21 ± 0.20
	Spinal cord	0.50 ± 0.16	0.87 ± 0.44	0.17 ± 0.26
Prostate	PTV	0.68 ± 0.15	0.74 ± 0.19	0.88 ± 0.24
	Bladder	0.39 ± 0.12	0.58 ± 0.19	0.05 ± 0.04
	Rectum	0.87 ± 0.49	0.98 ± 0.25	0.27 ± 0.16

Y. Katsuta *et al. Physica Medica* 42 (2017) 1–6



Summary of AI for IMRT QA

Prediction of IMRT QA results (Γ -index, dose diff., etc.)

- **Input data:** Plan attributes (complexity measures, field & fluence descriptors, beam energy, MU, etc.)
- **Training data:** *DICOM plans* and *IMRT QA results* from prior patients
- **Advantages:**
 - Directly applicable to varied IMRT QA technologies & strategies.
 - Can enhance current IMRT QA as early warning for likely failures
- **Disadvantages:** Subject to same limitations of QA device (inherent susceptibility of device to false positives/negatives) & analysis method (lack of relevance for Γ -index, etc.)

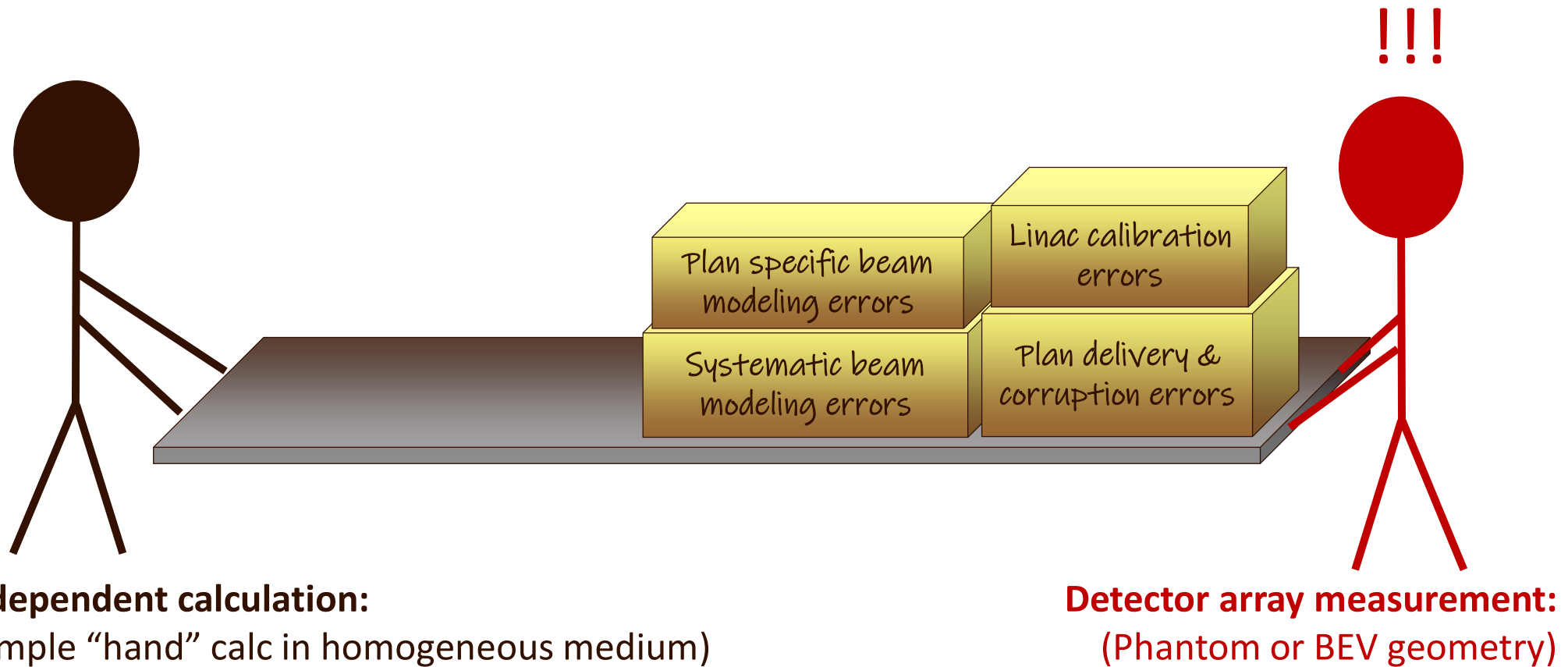
Prediction of machine parameters at treatment delivery

- **Input data:** Control point information from DICOM-RT plan (MLC pos. & velocity, gantry, etc.)
- **Training data:** *DICOM plans* and *trajectory files* from prior patients
- **Advantages:**
 - Direct prediction of relevant clinical measures (patient DVH).
 - Large dataset (1000s data-points per field)
 - Low maintenance training (no need for QA measurements-> can use clinical log files)
 - Model can be dynamic (updated daily)
- **Disadvantages:** Subject to same limitations as log files (does not measure beam symmetry, calibration errors, etc.)



Evolving Roles of Patient Specific QA:

Traditional IMRT QA:



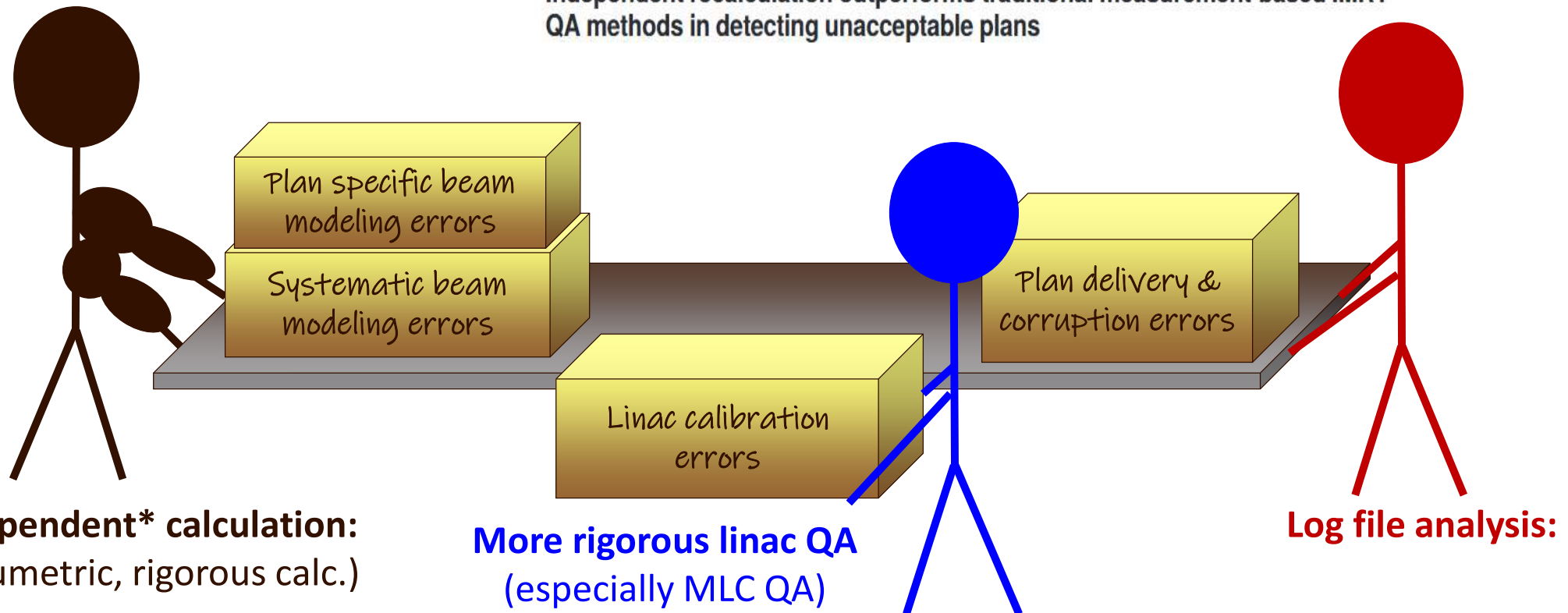


Evolving Roles of Patient Specific QA:

Log File Based IMRT QA:

Medical Physics, 46 (8), August 2019

Independent recalculation outperforms traditional measurement-based IMRT QA methods in detecting unacceptable plans



Esposito *et al. Rad. and Onc.*, Vol 149, pp 158-167, 2020.

*assuming independent commissioning (using separate beam data!) & unique beam model per TG-114, TG-219 recommendations



Broader Perspective & Looking Forward

Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee

Medical Physics, Vol. 30, No. 8, August 2003

ic's needs and may change with experience. However, it is important to emphasize that new users will need to spend much more time validating IMRT plans than is common with conventional treatments. Direct measurements will be necessary until independent dose calculation methods are developed and validated.

Independent recalculation outperforms traditional measurement-based IMRT QA methods in detecting unacceptable plans

Med. Phys. 46 (8), August 2019

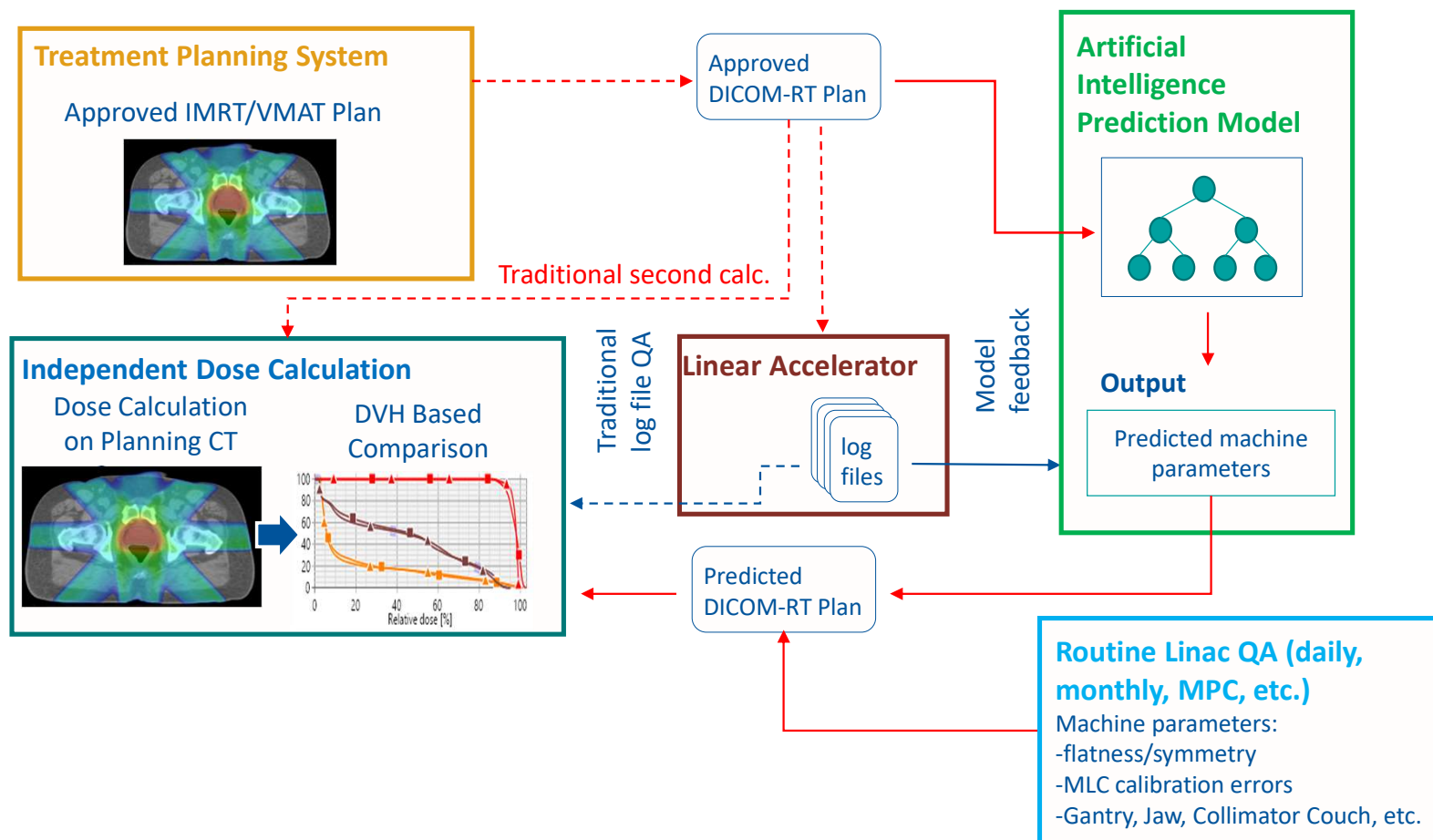
Results: Overall, based on common clinical criteria, the independent recalculation was 12 times more sensitive at detecting unacceptable (failing) IROC phantom results than clinical measurement-based IMRT QA. The recalculation was superior, in head-to-head comparison, to the EPID, ArcCheck, and MapCheck devices. The superiority of the recalculation vs these array-based measurements persisted under ROC analysis as the recalculation curve had a greater area under it and was always above that for these measurement devices. For detecting modest errors (poor phantom results rather than failing phantom results), neither the recalculation nor measurement-based IMRT QA performed well.

Conclusions: A simple recalculation outperformed current measurement-based IMRT QA methods at detecting unacceptable plans. These findings highlight the value of an independent recalculation, and raise further questions about the current standard of measurement-based IMRT QA. © 2019

- Best role of rigorous independent calc. + virtual QA / DVH analysis using prediction model?
- Potential to re-assess roles of:
 - Routine (linac) QA
 - Why verify linac calibration errors on a per patient basis???
 - Plan robustness checks
 - Plan quality & deliverability checks
 - Modified routine QA frequency
 - Machine QA
 - End-to-end tests
- FMEA will likely play an important role in re-structuring patient specific QA

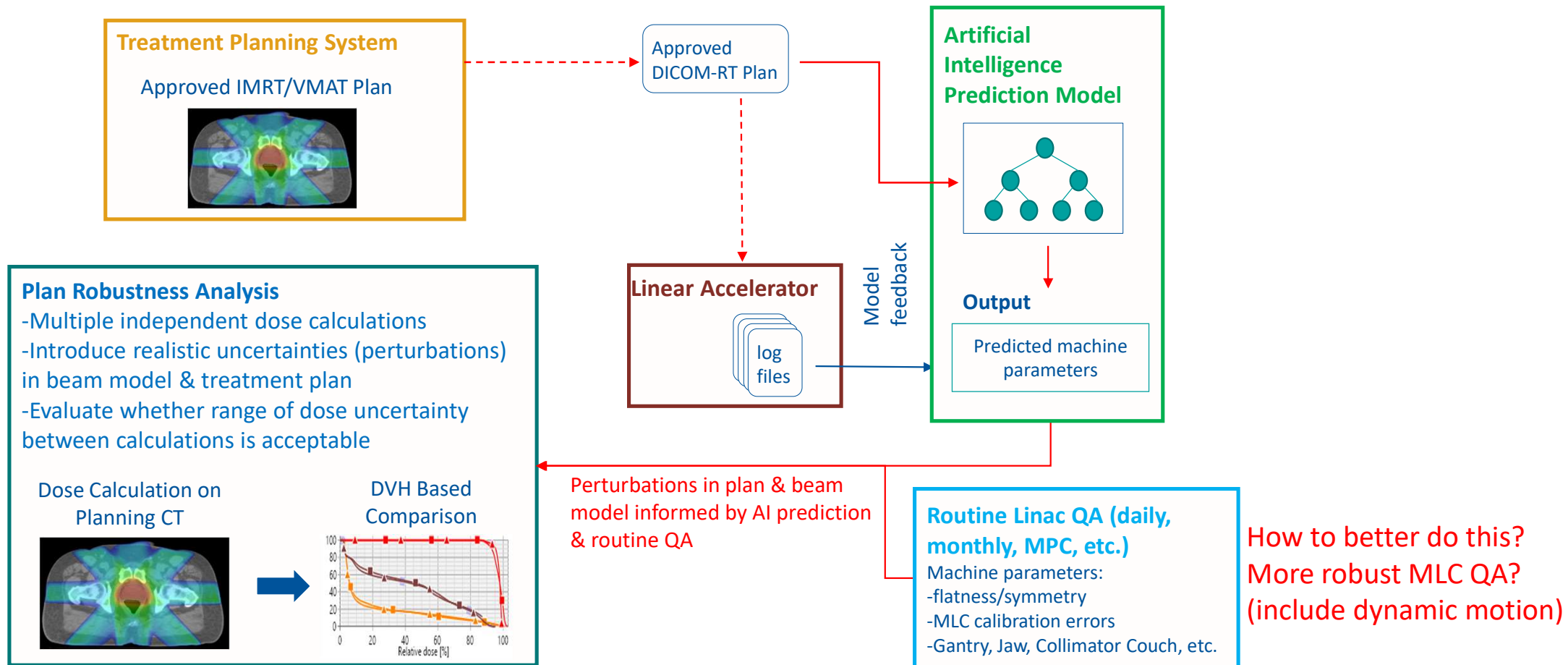


Incorporating Routine Linac QA?





Incorporating Plan Robustness Analysis?





Summary

- AI prediction models can be used to predict either results from IMRT QA measurements (Γ -index, dose diff., etc.) or machine parameters at treatment delivery (patient dose & DVH) with unique cost/benefit to each approach
- Opportunity to re-assess patient specific pre-treatment QA paradigm & the roles of various components:
 - Routine (linac) QA, plan robustness checks, plan quality & deliverability checks, use of routine end-to-end tests

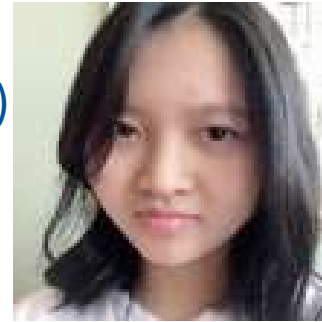


Questions?

Collaborators:

Duke University

- Brett Erickson (graduate student, Med. Phs. resident)
- Lam Lay (graduate student)
- Yunfeng Cui, PhD (faculty)
- William Giles, PhD (staff physicist)
- Jennifer O'Daniel, PhD (faculty)



Duke Kunshan University

- Kai-Cheng Chuang (graduate student)
- Yuqing Cui (graduate student)

